

SUPPLEMENTAL MATERIAL

The identification and exclusion of influential observations is an important aspect of our multivariable-adjusted logistic regression analyses. We have used the PROC SURVEYLOGISTIC procedure (SAS version 9.1) to estimate the prevalence odds ratios of periodontal disease for a three-fold increment in creatinine-corrected urinary cadmium concentrations. To identify those observations that exerted a large influence on the estimated parameter for the cadmium variable ($\log_e[\text{creatinine-corrected urine cadmium}]$) in our multivariable-adjusted logistic regression model we used DFBETA (or $\Delta\beta$) -- a diagnostic statistic that examines the effect of excluding a subject (or group of subjects) on the estimated coefficients (Hosmer and Lemeshow 2000). To obtain DFBETA we used PROC LOGISTIC, incorporating the appropriate survey weights. We subsequently identified potentially influential observations using plots of DFBETA vs predicted probabilities, and re-estimated the OR for periodontal disease removing one potential outlier at a time. We considered any observation as an influential point if excluding it brought about a change of at least 20% in the $\log_e[\text{OR}]$ odds of periodontal disease. In this manner, we identified 5 influential points in our whole study population. All of these participants were cases and represented 0.8% of all periodontal disease sufferers included in our multivariate analyses. Their creatinine-corrected urine cadmium concentration ranged from 0.01 to 1.23 $\mu\text{g/g}$. Without removing these five observations, the OR for periodontal disease for a three-fold increase in urine cadmium concentrations was 1.32 (95% CI: 1.06 – 1.64). Exclusion of these influential points changed the OR to 1.54 (95% CI: 1.26 – 1.87).

In a similar manner, in the sub-group with limited tobacco exposure, we identified six influential observations. Removal of these influential points resulted in a substantial change in the association between urinary cadmium and periodontal disease. When these six observations were retained, the OR for periodontal disease for a three-fold increase in urine cadmium was 1.05 (95% CI: 0.67 – 1.64). On excluding the six observations, the OR increased to 1.68 (95% CI: 1.26 – 2.24). Notably, further exclusion of observations did not produce an appreciable change in the parameter estimate for the cadmium variable. For example, deletion of two observations with the next highest DFBETA scores changed the OR to 1.64 (95% CI: 1.25 – 2.16).

We examined the characteristics of the six influential observations to identify potential reasons for such a large effect on the association between cadmium and periodontal disease (see Table). Notably, all six participants had periodontal disease and three of the six participants had cadmium concentrations markedly lower than the geometric mean urine cadmium concentration of the low-tobacco group (0.22 µg/g creatinine). While these low-cadmium-exposed cases represent a small fraction of all periodontal disease sufferers in this sub-group (approximately 8%), their inclusion exerts a substantial influence on the association between urinary cadmium and periodontal disease. We urge caution in interpreting the results in the low-tobacco participants as the statistically significant association between cadmium and periodontal disease is reliant on the removal of these influential observations.

Table. Characteristics of six low-tobacco exposed participants identified as influential observations.

Participant	Urine cadmium concentration ($\mu\text{g/g}$ creatinine)	Periodontal disease	Age (years)	Sex
1	0.01	Yes	50	Female
2	0.20	Yes	36	Female
3	0.21	Yes	60	Female
4	0.04	Yes	37	Male
5	0.03	Yes	52	Male
6	3.63	Yes	54	Female

REFERENCE

Hosmer DW, Lemeshow S. 2000. Assessing the fit of the model. In: Applied logistic regression (2nd ed). New York: Wiley, 167-188.